AWARD NUMBER: W81XWH-12-2-0052

TITLE: Integrative Cardiac Health Project, Windber Research Institute

PRINCIPAL INVESTIGATOR: Darrell L. Ellsworth, PhD

RECIPIENT: Windber Research Institute
Windber, PA 15963

REPORT DATE: July 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2015	Annual	22Jun2014 - 21Jun2015

4. TITLE AND SUBTITLE
Integrative Cardiac Health Project, Windber Research Institute

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-2-0052

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Darrell L. Ellsworth, PhD

Email: d.ellsworth@wriwindber.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Windber Research Institute
Windber, PA 15963

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

There is an increasing prevalence of obesity and cardiovascular disease (CVD) risk factors in the military population, which is negatively affecting operational readiness. The ability to prevent heart disease and reduce its overall impact on morbidity, would increase the quality of life among military personnel and their dependents, and has the potential to generate enormous cost savings for the DoD. In the Integrative Cardiac Health Program (ICHP), we are investigating physiological and molecular responses to risk factor modification interventions in individuals or populations at risk for cardiovascular disease (CVD). We aim to better understand CVD risk at the molecular level before onset of clinical disease, and develop outcomes-based patient empowering lifestyle solutions to prevent disease. Through this research, our objectives are to (1) identify genetic influences on CVD and integrate information on dietary, behavioral, and lifestyle factors to provide important information on CVD risk reduction and (2) discover new genes in previously associated pathways to reveal new molecular influences on cardiovascular risk reduction.

15. SUBJECT TERMS- Nothing listed

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	33	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

1. **INTRODUCTION:** There is an increasing prevalence of obesity and cardiovascular disease (CVD) risk factors in the military population, which is negatively affecting operational readiness. The ability to prevent heart disease and reduce its overall impact on morbidity, would increase the quality of life among military personnel and their dependents, and has the potential to generate enormous cost savings for the DoD. In the Integrative Cardiac Health Program (ICHP), we are investigating physiological and molecular responses to risk factor modification interventions in individuals or populations at risk for cardiovascular disease (CVD). We aim to better understand CVD risk at the molecular level before onset of clinical disease, and develop outcomes-based patient empowering lifestyle solutions to prevent disease. Through this research, our objectives are to (1) identify genetic influences on CVD and integrate information on dietary, behavioral, and lifestyle factors to provide important information on CVD risk reduction and (2) discover new genes in previously associated pathways to reveal new molecular influences on cardiovascular risk reduction.

2. **KEYWORDS:** Lifestyle modification, cardiovascular disease, obesity, gene expression, RNA sequencing, gender differences, molecular response, diet, exercise.

3. **OVERALL PROJECT SUMMARY:**
   For all tasks, our efforts were devoted to generating as much high quality DNA and RNA sequence data as possible. We made progress in analyzing DNA methylation changes in response to cardiovascular risk reduction: DNA was isolated from 102 whole blood samples, Methyl-Mini Sequencing was completed on 60 samples, 30 samples are currently undergoing sequencing, and 30 additional samples will be sequenced as soon as possible.

   The first large-scale RNA sequencing run using 40 samples yielded 123.579 billion bases of sequence, with an average of 3.09 billion bases per sample, and 88.5% of the reads having a quality score >Q30. (06/22/2014—09/21/2014)

   The second paired-end 200 cycle sequencing run yielded 1.00817 billion reads over the entire flow cell, which generated ~187 billion bases of sequence and an average of 126 million reads per flow cell lane; on average 81.45% of the reads PF were $\geq$ Q30 (99.9% accurate). (09/22/2014—12/21/2014)

   A third paired-end 200 cycle sequencing run yielded 2.088 billion reads and generated ~388 billion bases sequenced. Each lane produced an average of 260.96 reads, and 91% of those reads passed the HiSeq filter screen. (12/22/2014—03/21/2015)

   A fourth paired-end 200 cycle sequencing run produced on average 237 million reads; 86% of the reads had a Q-score >30. (06/1/2015—6/21/2015)

Continuing review for all research protocols for this project were submitted to and approved by the Chesapeake IRB. (06/22/2014—6/21/2015).

Task #1: Epigenetic changes in DNA (genome-wide patterns of methylation) during CV risk reduction

In this task, we are examining patterns of DNA methylation across the entire genome in circulating leukocytes in response to cardiovascular risk reduction (lifestyle and surgically-assisted) using new and current participants in our ICHP programs. We will seek to identify changes in methylation in specific areas of the genome and relate these changes to known and novel genes influencing heart disease. Results from this research may be useful in further understanding molecular mechanisms associated with changes in CV risk factors and regulatory processes involved in heart disease development.

During the year, DNA was isolated from 102 whole blood samples (Table 1) using the Quick gDNA Blood Mini kit (Zymo Research) in the following groups: intensive lifestyle baseline (n=23) and one year (n=23), laparoscopically placed adjustable gastric banding (LAGB) baseline (n=23) and one year (n=23), and control baseline (n=5) and one year (n=5) as shown in the table below. DNA concentrations were 60.1±36.4 ng/μl (range 14.1-269.8 ng/μl), OD260/280 ratios were 2.01±0.10 (range 1.78-2.60) and OD260/230 ratios were 1.79±0.96 (range 0.13-6.31).

Table 1. Concentrations and purity measures for DNA isolated from whole blood from laparoscopically placed adjustable gastric banding patients.

Patient ID	Time point	DNA Concentration (ng/μl)	Abs 260/280	Abs 260/230
BATCH #2				
05-03-047.2	Baseline	33.4	1.85	0.93
05-03-047.3	1 year	14.1	2.60	3.69
05-03-154.2	Baseline	65.6	1.97	2.11
05-03-154	1 year	66.1	1.99	0.92
05-03-174.2	Baseline	72.3	2.00	2.43
05-03-174.2	1 year	76.8	2.02	1.64
05-03-185.3	Baseline	31.2	2.08	1.31
05-03-185	1 year	78.3	1.96	2.14
05-03-166.2	Baseline	63.8	2.12	2.44
05-03-166.2	1 Year	85.1	2.00	2.00
cv000303.2	Baseline	22.1	2.06	---
cv000303.2	1 year	35.8	2.09	1.50
cv000589.2	Baseline	102.5	2.06	2.35

cv000589.2	1 year	56.4	2.19	4.27
cv000301	Baseline	49.1	2.07	6.31
cv000301.2	1 year	39.9	2.03	1.42
cv000812	Baseline	59.8	1.90	3.31
cv000812.2	1 year	61.4	2.15	4.81
cv000919.2	Baseline	29.8	2.01	0.45
cv000919	1 year	36.7	1.85	4.41
cv000613	Baseline	62.5	1.90	2.7
cv000613	1 year	35.4	2.04	3
cv000377	Baseline	24.0	2.09	---
cv000377.2	1 year	42.1	1.92	---
cv000709.2	Baseline	38.8	1.93	1.62
cv000709.2	1 year	35.3	1.83	0.55
cv000743.2	Baseline	43.1	2.00	0.18
cv000743.2	1 year	35.5	1.88	0.46
cv000226	Baseline	39.3	1.89	0.88
cv000226.2	1 year	35.3	1.78	---
BATCH #3				
05-03-003	Baseline	67.7	2.03	1.04
05-03-003.2	1 Year	64.1	2.10	2.15
05-03-148.2	Baseline	125.4	1.99	2.16
05-03-148	1 year	23.3	2.10	1.47
05-03-163.2	Baseline	31.2	1.98	2.26
05-03-163.2	1 Year	80.6	2.14	2.08
05-03-177.2	Baseline	65	1.99	0.80
05-03-177	1 year	62.5	2.11	1.63
05-03-179	Baseline	134.8	1.99	2.26
05-03-179	1 Year	113.7	1.99	2.14
05-03-193.2	Baseline	40.4	2.01	0.24
05-03-193.2	1 year	79.2	2.04	2.01
05-03-194	Baseline	75.6	2.06	1.80
05-03-194	1 Year	77.9	2.05	1.85
05-03-198.2	Baseline	150.3	1.96	2.09
05-03-198	1 year	51.6	2.14	2.29
05-03-157	Baseline	80.8	2.05	2.45
05-03-157	1 year	88.5	2.06	1.25
cv287	Baseline	41.0	2.12	2.19
cv287	1 year	60.0	2.06	1.57
cv288.2	Baseline	127.2	1.94	2.24
cv288.2	1 year	269.8	1.98	1.93
cv308	Baseline	53.6	2.07	2.06
cv308	1 year	29.1	2.13	1.94

cv350	Baseline	38.6	1.98	1.00
cv350	1 year	47.9	1.98	1.21
cv613 int.2	Baseline	36.6	2.11	1.15
cv613 int.2	1 year	32.1	2.16	1.83
cv749.2	Baseline	70.7	2.01	2.11
cv749	1 year	24.7	2.23	1.42
cv823	Baseline	145.9	2.05	2.00
cv823	1 year	60.5	2.05	1.88
cv884	Baseline	57.1	2.01	1.90
cv884.2	1 year	73.5	1.97	2.17
cv579	Baseline	76.3	2.01	1.69
cv579	1 year	49.7	1.96	1.70
BATCH #4				
05-03-004.2	Baseline	156.4	1.98	1.40
05-03-004	1 Year	71.0	1.91	1.23
05-03-035.2	Baseline	49.2	1.91	0.70
05-03-035	1 Year	19.1	1.87	1.63
05-03-162	Baseline	22.4	2.15	0.96
05-03-162	1 Year	86.7	1.98	1.02
05-03-176	Baseline	97.3	2.00	2.00
05-03-176	1 Year	110.6	2.03	2.03
05-03-178	Baseline	105.3	2.04	2.21
05-03-178	1 Year	83.4	2.04	2.25
05-03-186	Baseline	64.2	2.02	2.51
05-03-186.2	1 Year	52.7	2.04	3.07
05-03-211.2	Baseline	44.7	1.88	0.80
05-03-211.2	1 Year	46.2	1.99	1.75
05-03-225	Baseline	78.5	1.96	1.97
05-03-225	1 Year	72.3	1.96	1.71
05-03-181	Baseline	39.4	1.98	2.90
05-03-181	1 Year	50.9	1.98	1.15
427	Baseline	41.8	2.02	1.55
427	1 Year	34.1	1.92	1.20
428	Baseline	33.2	1.88	1.61
428	1 Year	35.1	1.84	0.91
432	Baseline	43.6	1.99	1.03
432	1 Year	49.9	2.00	1.39
669	Baseline	32.3	2.11	2.34
669	1 Year	52.0	1.93	1.44
687	Baseline	36.3	1.97	0.13
687	1 Year	32.3	1.93	0.52
882	Baseline	35.4	1.95	1.47

882	1 Year	27.7	1.96	2.10
918	Baseline	22.1	2.04	1.28
918	1 Year	27.4	2.03	1.72
928	Baseline	31.2	1.98	0.82
928	1 Year	55.0	1.93	0.93
434	Baseline	42.7	2.02	0.21
434	1 Year	69.4	2.00	2.01

Library construction was completed on all 30 samples in Batch #2 and all samples passed QC requirements. One microgram of DNA from each library was sequenced using Methyl-Mini Sequencing, which is a Reduced Repression Bisulfite Sequencing method that allows for detection of 3-4 million CpG sites throughout the genome. All samples had a bisulfite conversion rate of >98.25% (Table 2). The number of CpG (methylated) sites per sample was >7,500,000, with minimum coverage of 5X per sample. The top 2000 hypo-methylated (decreasing methylation) and hyper-methylated (increasing methylation) sites in the three groups were identified.

Table 2. Results of Reduced Repression Bisulfite Sequencing on 30 laparoscopically placed adjustable gastric banding patients.

Sample	No. Total Reads	No. Mapped Reads	Mapping Ratio	Unique CpG	BS Conv. Rate
047B	36,705,287	21,623,797	58.91%	8,380,942	98.95%
047Y	39,480,981	22,182,774	56.19%	8,449,711	98.85%
154B	39,344,051	22,515,910	57.23%	8,427,361	98.95%
154Y	46,518,809	22,654,277	48.70%	8,034,750	98.26%
174B	37,535,399	21,838,695	58.18%	8,380,190	98.84%
174Y	35,271,999	19,348,584	54.86%	8,107,389	98.78%
185B	43,264,229	23,560,392	54.46%	8,589,618	98.34%
185Y	29,840,305	16,384,027	54.91%	8,009,735	98.98%
166B	35,539,525	19,793,424	55.69%	8,284,017	98.92%
166Y	38,190,627	21,252,040	55.65%	8,356,105	99.06%
303B	35,952,416	20,634,723	57.39%	8,418,795	98.92%
303Y	40,202,579	23,604,582	58.71%	8,916,549	99.02%
589B	35,525,846	21,494,984	60.51%	8,765,446	98.97%
589Y	39,022,506	23,838,829	61.09%	9,056,597	98.83%
301B	31,184,095	18,317,063	58.74%	8,391,100	99.08%
301Y	32,292,061	17,936,325	55.54%	8,089,337	98.78%
812B	34,115,895	19,572,784	57.37%	8,133,044	99.02%
812Y	38,180,314	21,509,842	56.34%	8,235,083	98.92%

919B	41,441,318	23,416,147	56.50%	8,451,510	98.74%
919Y	35,784,427	21,502,356	60.09%	8,387,346	98.96%
613B-C	42,637,181	23,477,588	55.06%	8,391,188	98.76%
613Y-C	39,364,561	21,509,836	54.64%	8,267,203	98.86%
377B	39,764,053	22,564,502	56.75%	8,343,424	98.72%
377Y	39,329,753	22,362,178	56.86%	8,349,635	99.06%
709B	32,495,359	17,818,459	54.83%	7,872,673	98.57%
709Y	37,557,280	21,269,135	56.63%	8,098,635	98.35%
743B	40,901,341	23,332,313	57.05%	8,448,724	98.69%
743Y	39,784,766	23,102,904	58.07%	8,366,518	99.28%
226B	34,963,505	20,175,862	57.71%	8,254,083	98.74%
226Y	37,830,893	22,011,047	58.18%	8,367,000	99.06%

For LAGB patients, a heat map based on differences in methylation between baseline and one year (left) and a pairwise scatter plot of methylation patterns at baseline and one year (right) are shown below in Figure 1.

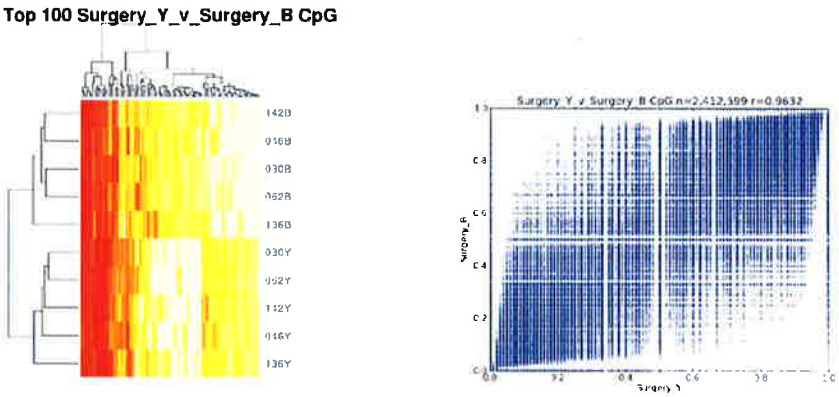

Figure 1. Differences in methylation between pre-surgery and one year post-surgery in 30 LAGB patients depicted by a heat map (left panel) and pairwise scatter plot (right panel).

DNA samples from Batch #3 were sent to Zymo Research for Methyl-Mini Sequencing. Results are anticipated by August 2015. DNA samples from Batch #4 will be sent to Zymo Research during the next quarter.

Task #2: Profile metabolic activity in blood and adipose tissue during surgical weight loss

During the year, no new patients undergoing laparoscopically placed adjustable gastric banding (LAGB) were enrolled in the study. No additional follow-up blood samples or adipose tissue samples were collected. Total RNA was isolated from 79 PAXgene peripheral blood samples from 52 patients (Table 3). RNA concentrations were 77.1±40.8 ng/μl (range 12.4-181.5 ng/μl), OD260/280 ratios were 2.17±0.09

(range 1.87-2.53), and RIN numbers were 8.18±0.41 (range 6.8-9.1). Forty-three RNA samples were run on Affymetrix gene expression arrays with call rates of 59.99±1.51% (range 55.41-61.80%).

A summary of time points for which RNA was isolated from peripheral blood during the year is as follows: baseline pre-surgery (n=14), five to seven months post-surgery (n=24), one year post-surgery (n=10), one year one month to one year eleven months (n=13), two years to two years eleven months (n=2), three years six months (n=1), four years six months (n=1), five years (n=3), five years one month to five years eleven months (n=7), and six years post-surgery (n=4).

Table 3. Concentrations, purity measures, and call rates on gene expression arrays for RNA isolated from whole blood from laparoscopically placed adjustable gastric banding patients.

Sample	Time Point	Concentration (ng/μl)	OD260/280	RIN	Call Rate (%)
05-03-001	6 YR	53.33	2.10	8.2	
05-03-005	6 YR	76.35	2.11	7.8	
05-03-031	5 YR 10 MO	22.97	1.87	8.7	61.35
05-03-049	5 YR	135.71	2.22	7.7	58.01
05-03-050	5 YR 8 MO	69.06	2.11	8.8	
05-03-050	6 YR	73.73	2.10	8.7	
05-03-050	5 YR 8 MO	69.06	2.11	8.8	
05-03-053	5 YR 6 MO	172.84	2.14	7.7	
05-03-053	6 YR	63.09	2.10	8.2	
05-03-053	5 YR 6 MO	172.84	2.14	8.0	
05-03-073	5 YR 10 MO	44.15	2.17	8.2	
05-03-077	5YR 2 MO	42.45	2.34	8.1	60.53
05-03-084	5 YR	68.90	2.12	7.8	
05-03-091	BASE	113.52	2.20	8.4	60.09
05-03-094	5 YR	43.33	2.09	8.5	
05-03-097	BASE	33.91	2.06	7.7	57.54
05-03-112	BASE	106.94	2.15	7.6	58.23
05-03-115	4 YR 6 MO	25.39	2.19	8.0	60.48
05-03-126	BASE	94.63	2.11	7.9	
05-03-126	3 YR 6 MO	29.55	2.25	8.6	
05-03-136	1 YR 7 MO	56.26	2.33	7.9	60.23
05-03-157	2 YR 4 MO	175.34	2.17	8.7	59.39
05-03-163	6 MO	37.41	2.29	7.9	60.50
05-03-166	2 YR	37.08	2.10	8.3	61.50
05-03-167	7 MO	70.29	2.10	8.1	

05-03-167	BASE	64.83	2.25	6.8	55.41
05-03-169	1 YR 8 MO	65.40	2.10	8.3	
05-03-173	1 YR 6 MO	76.49	2.16	8.4	61.38
05-03-175	1 YR 6 MO	144.48	2.15	9.1	59.82
05-03-175	6 MO	143.33	2.20	8.4	60.00
05-03-181	1 YR 6 MO	50.91	2.22	7.9	60.23
05-03-182	6 MO	56.59	2.10	8.0	59.88
05-03-185	1 YR 6 MO	62.80	2.17	8.1	
05-03-192	1 YR	40.33	2.24	9.1	61.74
05-03-192	6 MO	40.58	2.53	7.7	60.25
05-03-192	1 YR 6 MO	93.90	2.15	8.3	
05-03-194	6 MO	104.74	2.17	8.2	57.71
05-03-195	BASE	81.92	2.12	8.0	60.09
05-03-196	7 MO	145.65	2.17	8.9	56.61
05-03-197	7 MO	112.88	2.19	8.2	61.50
05-03-198	1 YR 6 MO	85.09	2.31	7.7	60.51
05-03-199	6 MO	45.84	2.04	8.7	61.38
05-03-204	7 MO	158.56	2.12	8.7	59.07
05-03-208	1 YR 6 MO	77.65	2.18	8.6	61.80
05-03-210	1 YR 5 MO	76.97	2.11	7.8	
05-03-210	5 MO	111.05	2.19	7.7	60.04
05-03-210	5 MO	111.05	2.19	7.7	
05-03-211	1 YR	12.42	2.16	8.7	61.65
05-03-212	1 YR	89.43	2.11	8.2	60.52
05-03-212	5 MO	114.51	2.13	8.0	
05-03-212	BASE	176.70	2.23	7.4	58.63
05-03-214	6 MO	63.75	2.16	8.4	
05-03-214	1 YR 3 MO	60.66	2.12	8.7	
05-03-215	6 MO	181.52	2.14	8.0	
05-03-216	6 MO	79.05	2.19	8.6	61.80
05-03-218	6 MO	44.07	2.30	7.8	61.65
05-03-222	1 YR 4 MO	80.87	2.15	8.4	
05-03-222	6 MO	37.33	2.09	7.8	61.59
05-03-225	BASE	72.82	2.14	7.8	58.57
05-03-225	1 YR	21.95	2.48	8.1	61.54
05-03-227	7 MO	64.34	2.17	8.4	
05-03-227	BASE	74.35	2.33	7.9	59.28
05-03-228	6 MO	41.57	2.19	8.2	
05-03-228	1 YR	35.40	2.07	8.5	
05-03-228	BASE	61.44	2.23	7.7	
05-03-229	6 MO	23.92	2.00	8.3	59.03
05-03-229	1 YR	129.28	2.13	8.3	
05-03-229	BASE	40.06	2.23	8.1	61.73

05-03-232	6 MO	111.35	2.28	8.0	59.60
05-03-236	1 YR 1 MO	58.69	2.09	7.8	
05-03-236	6 MO	68.59	2.11	7.8	
05-03-236	1 YR	58.69	2.09	8.3	
05-03-244	1 YR	29.95	2.07	8.5	
05-03-244	BASE	66.80	2.13	7.8	59.74
05-03-244	6 MO	71.49	2.14	8.3	
05-03-253	1 YR	65.01	2.10	8.5	60.63
05-03-253	BASE	81.66	2.22	8.2	58.16
05-03-258	BASE	66.59	2.11	8.3	
05-03-258	1 YR	67.80	2.22	8.6	

Large-scale RNA Sequencing
During the year, supply issues continued to prevent us from being in full production mode for generating RNA sequence data on the HiSeq machine.

We began working with the new Illumina TruSeq Stranded Total RNA Library preparation kit by optimizing and streamlining the protocol. We found that this new stranded total RNA prep yields a broader range of transcripts and allows for detection of more precise information. First, we created 12 RNA libraries from 50 ng of previously globin-cleared total RNA, which was isolated from blood of obese patients via PAXgene. All libraries were quality checked via flurometrics and the bioanalyzer. Quantitative analysis showed good average concentration (35 ng/μl) per library. The bioanalyzer showed a nice well defined peak in the 270 bp range for each library and an average size distribution of ~290 bp which is perfect for our RNAseq needs on the Illumina HiSeq platform (Figure 2).

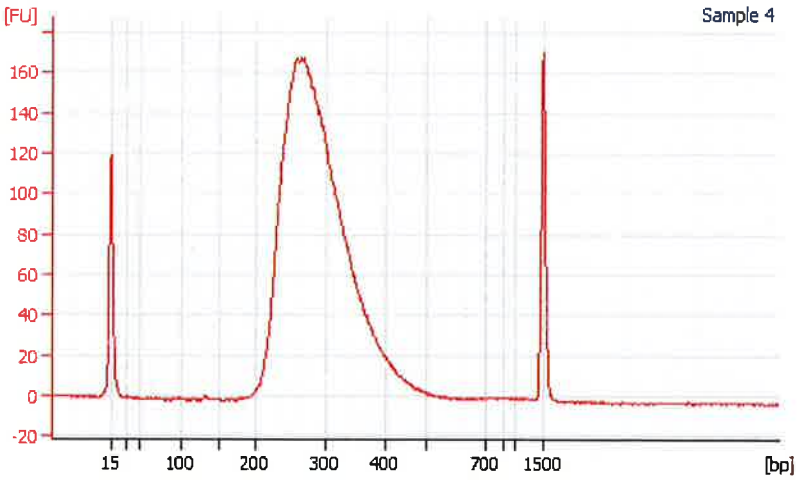

Figure 2. Bioanalyzer trace showing fragment size distribution for an RNA library to be used for RNA sequencing.

In the first run, large-scale RNA sequencing was conducted on 40 samples.

Data Analysis: Quality control (QC) and analysis work flow of the HiSeq data consisted of converting the generated BCL image files to FASTQ sequence files. Once the FASTQ files were generated, the data were de-multiplexed to separate the data for each individual sample. The individual data showed that overall, the run consisted of 1.02 billion reads, with an average 25.75 million reads per sample: >95% of these reads passed the HiSeq quality filter. In terms of RNA sequence, the run yielded 123.579 billion bases of sequence, with an average of 3.09 billion bases per sample. Clustering distribution per lane was very good, with 5 samples per lane the clustering percent across the entire flow cell was around 19%, with 20% being perfect. For the entire run, 88.5% of the reads had a quality score >Q30, with the average mean for all reads being Q35 (Q30 is defined as 99.9% accuracy).

After demultiplexing was complete the individual data sets from each sample were further checked to assess quality of particular samples by looking at base content per cycle along with adapter content, duplication levels, and Kmer content (Figure 3).

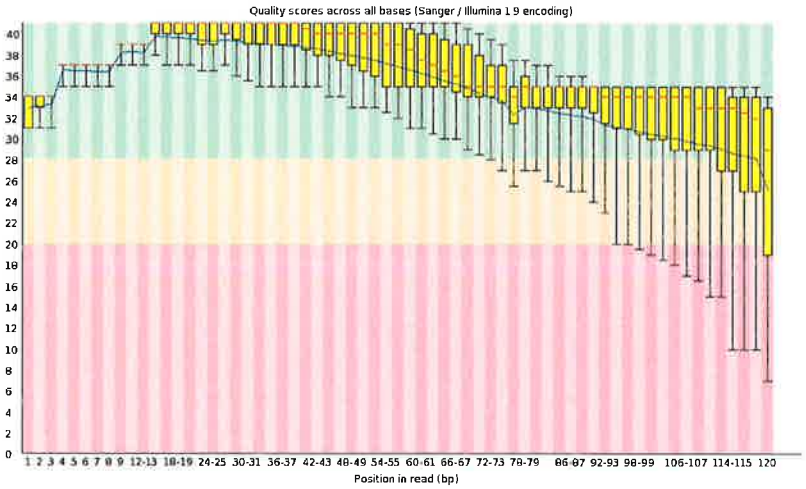

Figure 3. Base content per cycle and associated quality scores obtained from RNA sequencing.

Next, using the information from the last QC, the data was trimmed and cleaned of low quality sequences. Before trimming/clean-up there were a total of 1,029,826,359 total reads, after trimming/clean-up we were left with 1,020,546,890 reads which is 99.1% of the total reads. After trimming was complete the reads were mapped to the human genome (hg19); on average 83% of the reads per sample were mapped to the human genome. This translates into an average of 21.2 million mapped reads per sample which is good for RNAseq Gene expression analysis.

The RNA-Seq reads for each gene were then counted, which yielded a total of 63,677 genes identified; 60% of the genes identified were protein coding genes and pseudogenes. Of the 63,677 genes identified, 13,218 of them had sufficient reads to

provide statistically significant information, 90% of these 13,218 genes are protein coding genes and pseudogenes.

To conduct a second large-scale sequencing run, we performed qPCR on 12 previously generated total RNA libraries using the KAPA Biosystems library quantification kit. All libraries had good concentrations (average of 185.6 nM) and were suitable for sequencing. In addition, 16 new total RNA libraries were created and tested for quality and quantity using the Agilent Bioanalyzer and Life Technologies Qubit. All libraries were of good fragment length (~280 bp) and showed good concentrations (average=32.6 ng/μl). We then tested cluster efficiency and optimized data output by loading a flow cell at varying concentrations (15, 17, 18, 19 pM) in duplicate across the 8 lanes. After clustering was complete on the Cbot clustering station, the clustered flow cell was loaded onto the HiSeq machine and a 50-cycle single-end-read run was performed. The results showed that our libraries clustered at a density ranging from 1100 k/mm^2 to 1170 k/mm^2 and cluster density peaked at an input of 17 pM – higher input concentrations showed a drop in quality with little to no increase in cluster density. In a second assay, libraries of varying concentrations (12.5, 13, 13.5, 14 pM) were loaded on a HiSeq flow cell and a second 50-cycle single-read run was performed. All lanes performed well with cluster densities ranging from 990 k/mm^2 through 1116 k/mm^2. The test showed that the quantity of clusters is directly related to the quality of the data produced by the run. The 12.5 pM input resulted in a cluster density of 990 k/mm^2 where 89.1% of clusters Passed Filter (PF) and 96% of the reads were ≥ quality score Q30; whereas, 14 pM input resulted in cluster density of 1061 k/mm^2 where 84.8% of clusters PF and 94.8% of the reads were ≥ Q30. The denser the clusters the more difficult it was for the machine to differentiate individual clusters, which resulted in a drop in quality (see Table below).

Table 4. Input concentrations, cluster densities, and quality scores for the first RNA sequencing run on the HiSeq 2000.

Lane	Input Concentration	Density (K/mm2)	Clusters PF (%)	Reads (M)	Reads PF (M)	% ≥ Q30
1	12.5 pM	990 +/- 88	89.1 +/- 3.0	273.77	243.29	96.0
2	13 pM	991 +/- 89	88.8 +/- 3.4	274.11	242.74	95.8
3	13.5 pM	1041 +/- 75	86.6 +/- 4.1	287.74	248.67	95.2
4	14 pM	1061 +/- 75	84.8 +/- 6.5	293.23	247.85	94.8
5	12.5 pM	902 +/- 387	69.2 +/- 31.4	249.41	204.97	94.5
6	13 pM	1103 +/- 53	80.6 +/- 5.6	304.88	245.47	94.0
7	13.5 pM	1109 +/- 58	80.1 +/- 6.4	306.70	245.22	93.7
8	14 pM	1116 +/- 57	76.5 +/- 10.0	308.53	235.78	93.0

A HiSeq run was performed on 24 of the recently created total RNA libraries using a paired-end-read flow-cell with a 200 cycle reagent kit. Libraries were pooled into groups of 6 samples with one library pool to be run in each of the first 4 lanes of the flow cell and then duplicate pools run in the remaining lanes 5-8. After normalization and pooling, the multiplexed RNA library was diluted to the input clustering concentration of 13 pM, which was chosen based on the cluster density performance tests.

The paired-end 200 cycle sequencing run yielded 1.00817 billion reads over the entire flow cell, which generated ~187 billion bases of sequence and an average of 126 million reads per flow cell lane. 909.32 million reads (~90% of the 1.00817 billion reads) passed the HiSeq filter screen. On average 81.45% of the reads PF were $\geq$ Q30 (99.9% accurate) and 99.25% of the PF reads were identified and linked back to a particular sample. On average each sample had 37.6 million reads that passed filter. The average cluster density of the libraries was 911.5 k/mm^2.

A second HiSeq run was performed with a slight change to the cluster concentration input and using 24 of the recently created total RNA libraries. Libraries were pooled and run in the same orientation as the previous run. After normalization and pooling, the multiplexed library was diluted to the input clustering concentration of 12.5 pM.

The paired-end 200 cycle sequencing run yielded 2.088 billion reads and generated ~388 billion bases sequenced. On average each lane produced 260.96 reads, and 91% of those reads passed the HiSeq filter screen. Of those 256 million PF reads, 90% had a quality score of $\geq$ Q30. Average cluster density for each lane of the flow cell was 944 k/mm^2. 99.3% of the PF reads were identified and linked to a particular sample. PF read counts ranged from 66 million to 88.6 million (average of 78.1 million reads per sample). The mean quality score for all the identified PF reads was Q35.3 (Figure 4).

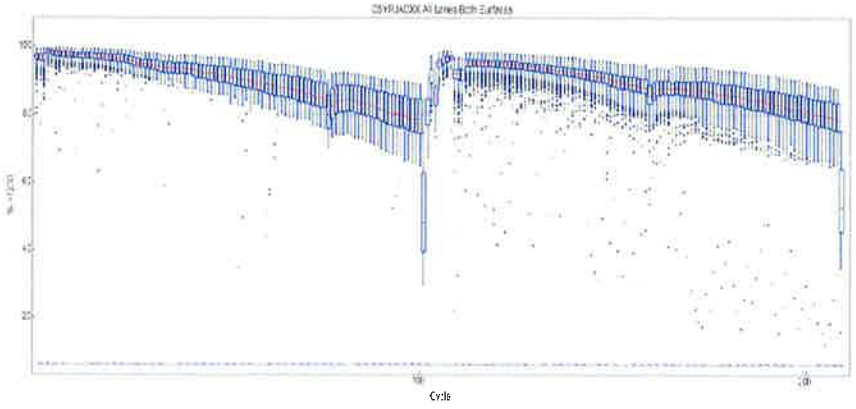

Figure 4. Quality scores per cycle for the first paired-end 200 cycle sequencing run.

Further analysis was conducted on the RNA sequencing data from these 2 HiSeq runs, which contained 24 samples each. The data were de-multiplexed and trimmed reads were aligned to the human genome. The runs were performed as paired end runs with different directional reads defined as left and right.

Run 1: On average each sample had 28 million left reads that mapped to the human genome out of a total of 37 million left reads (79% mapping rate for left reads). Percentage of left reads showing multiple alignments was 35% (range 14% to 77%). On average each sample had 11 million right reads that mapped to the human genome out of a total of 37 million right reads (33% mapping rate for right reads). Percentage of right reads showing multiple alignments was 32% (range 6% to 80%). The overall mapping rate for all reads was 56.34%; 32.96% of mapped reads showed multiple alignments. For all pairs of reads the average discordant alignments accounted for 11.38% of the mapped reads and the concordant pair mapping rate was 29.12%. The values for Run 1 were low due to leaks in the seals on the valves in the fluidics lines.

Run 2: On average each sample had 36 million left reads that mapped to the human genome out of a total of 38 million left reads (93% mapping rate for left reads). Percentage of left reads showing multiple alignments was 33% (range 11% to 80%). On average each sample had 36 million right reads that mapped to the human genome out of a total of 38 million right reads (94% mapping rate for right reads). Percentage of right reads showing multiple alignments was 33% (range 11% to 80%). The overall mapping rate for all reads was 93.56%; 32.54% of mapped read showed multiple alignments. For all pairs of reads the average discordant alignments accounted for 2.43% of the mapped reads and the concordant pair mapping rate was 87.83%. Run 2 performed much better than Run 1.

Total stranded RNA library preparation

A group of 36 previously globin-reduced total RNA samples were selected based on available time points and patient matching criteria; quantity and quality were checked via a spectrophotometer. RNA concentrations averaged 22.2 ng/μl (range 0.34-83.0 ng/μl). A subset of 12 samples (6 patients with baseline and 1 year time points) was selected for RNA library preparation based on quality and quantity. An aliquot of RNA from each sample was normalized to 5 ng/μl and 50 ng of total RNA was used to construct total stranded libraries.

Quality and quantity of the resulting libraries was assessed on an Agilent Bioanalyzer and a Life Technologies Qubit fluorometer. Average fragment size averaged 324 bp (range 309 bp to 359 bp). Concentrations of the individual libraries averaged 87 ng/μl (range 33-221 ng/μl). All libraries had well defined traces on the bioanalyzer (Figure 5) with a well-defined peak in the desired fragment size range. Concentrations showed good yield that should work well for the clustering procedure.

A second subset of 12 samples was selected from the previously quantified 36 to construct an Illumina Total Stranded RNA library. Samples were normalized to 5 ng/μl

and 50 ng of total RNA from each sample was used in the library preparation. When complete, the 12 libraries were checked for quality and quantity. The Bioanalyzer results indicated an average fragment size of 357 bp (range 335-383 bp), appropriate for sequencing. Concentrations of the libraries assessed via quantitative fluorometrics averaged 56.7 ng/μl (range 28-90 ng/μl). All libraries showed a well-defined peak in the desired fragment size range and were suitable for clustering on a flow cell.

Clustering HiSeq paired-end flow cell

The 24 libraries were normalized to 4 nM to start the flow cell clustering procedure. The libraries were then pooled according to their indices to create 4 tubes with 6 samples each. Samples were denatured and diluted to a cluster concentration of 12 pM. The 4 pools were loaded in duplicate onto the clustering station to begin clustering, during which the samples were bound to the flow cell.

Run 24 samples 6-plex on HiSeq

After the flow cell finished clustering the HiSeq was prepped and prepared for a 200 cycle single indexed, paired-end sequencing run. The sequencing run finished without error and performance statistics for all 8 lanes are as follows (Table 5). Cluster density averaged 1,004,000 clusters/mm^2 (range 963,000-1,050,000 clusters/mm^2), which is a large number of clusters. The % of clusters passing image filter averaged 88.4% (range 86.4-90.4%); number or reads that PF averaged 244.4 million reads (range 239.22-250.86 million); percentage of reads with a quality score of ≥Q30 averaged 89.9 (range 88.3-92.1) (Figure 6, Figure 7). Following quality filtering, each lane produced on average of 17.86 gigabases of sequencing data.

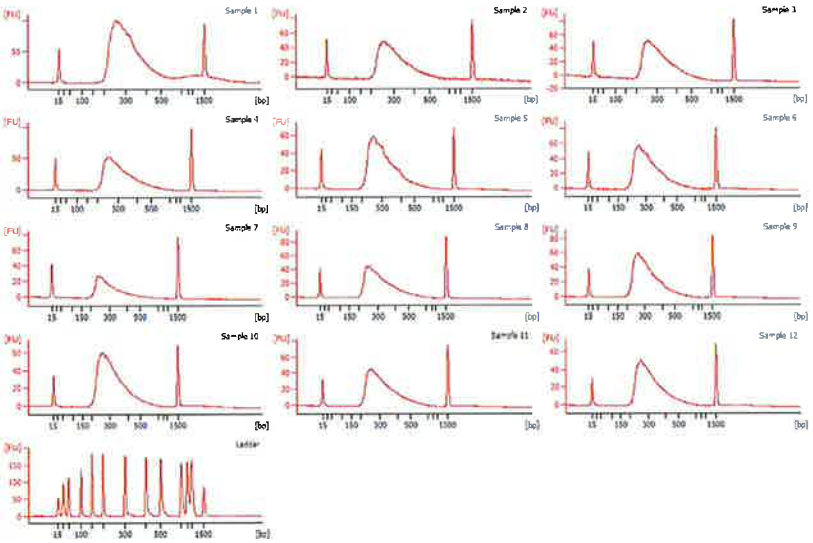

Figure 5. Bioanalyzer trace showing fragment size distribution for 12 RNA libraries used for RNA sequencing.

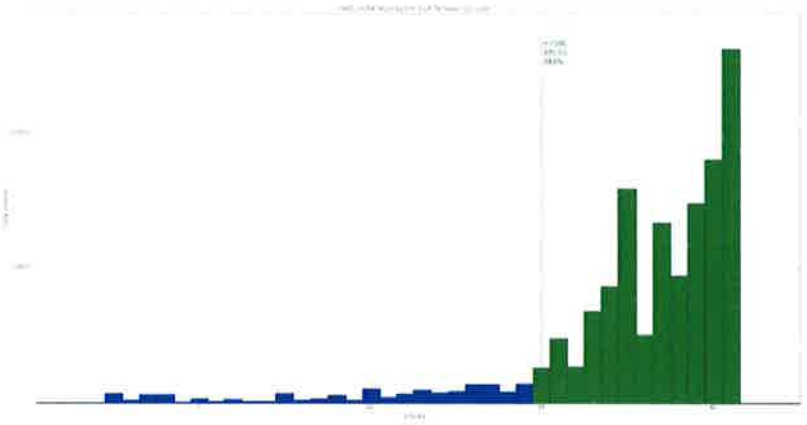

Figure 6. Bar graph showing the percentage of reads with a quality score of ≥Q30.

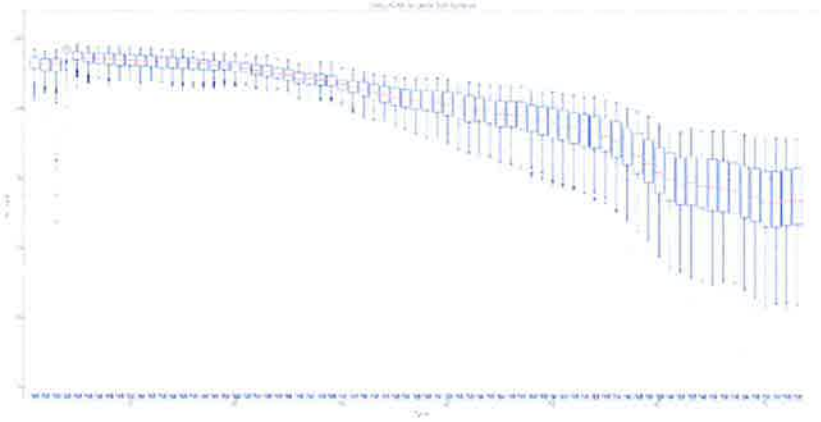

Figure 7. Quality scores per cycle for the second paired-end sequencing run.

Table 5. Cluster densities and quality scores for the second RNA sequencing run on the HiSeq 2000.

Cluster Density (k/mm^2)	Clusters PF	Reads (M)	Reads PF	% ≥Q30	Yield (G)
963	89.7	266.4	238.1	92.1	17.4
961	90.4	265.6	239.2	91.9	17.5
1045	86.9	288.9	250.2	88.6	18.3
1049	86.8	290.0	250.9	88.3	18.3
971	89.8	268.6	240.6	91.3	17.6
951	90.0	263.0	235.9	90.8	17.2
1050	86.4	290.3	250.1	87.4	18.3
1042	87.1	288.1	250.2	88.6	18.3

1004	88.4	277.6	244.4	89.9	17.9

Further analysis of this run showed that on average each of the HiSeq lanes had 99.3% of the PF reads identified and matched to a particular sample. Each lane of the flow cell contained 6 indexed libraries and the distributions of reads in each lane for each sample were very good at ±3.5% of the ideal distribution of 16.66% for each sample, with all but 4 samples being within ±1.5%.

Generate 24 New Total Stranded RNA-seq Libraries
Sixty RNA samples (Table 6) previously isolated from blood samples that were collected from obese patients in one of two intervention programs were located and checked for quality and quantity on the Bioanalyzer (Figure 8). Of the 60 samples, 24 samples (12 patients from our intervention programs, each with a baseline and 1 year time point) were selected for RNA-seq library preparation using the Illumina Total Stranded RNA Library preparation kit. Samples with very low concentrations or those with nonsense (negative) concentration values were treated as 0 and were not selected for RNA sequencing. These samples were used in their entirety for previous research. Any remaining RNA was below the level of detection.

Single indexed RNA-seq libraries were created from these 24 samples. Using the Agilent Bioanalyzer to check library quality and fragment size distribution, we determined that all libraries showed consistency as well as a well-defined product peak without unwanted contamination. Average fragment size was 310 bp (range 303-322 bp). Quantity analysis on the Qubit Fluorometer showed good product yield and good consistency in concentrations (average concentration was 44.5 ng/μl, range 33-56 ng/μl).

Cluster the 24 RNA-seq Libraries on a HiSeq Flowcell
The RNA libraries were all normalized to 4 nM and then pooled together in groups of 6 based on their indexes (Table 7). The 4 pools where then denatured, diluted, and clustered onto a paired end read flow cell at a concentration of 12 pM with one pool per lane through lanes 1, 2, 3, and 4 and then repeated in lanes 5 through 8.

Table 6. Concentrations and purity measures for RNA isolated from whole blood from lifestyle participants and laparoscopically placed adjustable gastric banding patients.

Sample ID	Date	ng/μl	260/280	260/230
Ornish BL 928	6/4/2015	-1.08	0.62	0.58
Ornish 1yr 928	6/4/2015	41.65	2.52	0.82
Ornish BL 307	6/4/2015	81.18	2.25	0.65
Ornish 1yr 307	6/4/2015	0.26	-0.12	0.08
Ornish BL 918	6/4/2015	3.99	9.54	20.53
Ornish 1yr 918	6/4/2015	7.27	8.88	1.87

Ornish BL 781	6/4/2015	105.94	2.31	0.28
Ornish 1yr 781	6/4/2015	9.81	4.53	0.24
Ornish BL 579	6/4/2015	62.77	2.26	0.88
Ornish 1yr 579	6/4/2015	79.43	2.22	0.9
Ornish BL 428	6/4/2015	75.61	2.26	1.29
Ornish 1yr 428	6/4/2015	76.54	2.17	1.5
Ornish BL 882	6/4/2015	45.94	2.35	0.63
Ornish 1yr 882	6/4/2015	42.89	2.28	0.34
Ornish BL 350	6/4/2015	85.14	2.2	1.2
Ornish 1yr 350	6/4/2015	78.85	2.3	1.08
Ornish BL 432	6/4/2015	65.23	2.19	1.55
Ornish 1yr 432	6/4/2015	41.88	2.35	1.59
Ornish BL 669	6/4/2015	6.11	28.08	0.58
Ornish 1yr 669	6/4/2015	-2.58	0.97	0.75
Ornish BL 434	6/4/2015	32.61	2.46	1.09
Ornish 1yr 434	6/4/2015	94.2	2.23	1.05
Ornish BL 287	6/4/2015	60.53	2.24	0.24
Ornish 1yr 287	6/4/2015	79.37	2.21	1.64
Ornish BL 687	6/4/2015	-2.38	0.78	0.6
Ornish 1yr 687	6/4/2015	4.66	3	1.4
Ornish BL 427	6/4/2015	48.38	2.42	1.32
Ornish 1yr 427	6/4/2015	71.02	2.36	0.6
Ornish 884 BL	6/29/2015	18.13	2.31	1.14
Ornish 884 1yr	6/29/2015	9.22	2.71	1.14
Marley 4 BL	6/23/2015	17.47	1.73	1.72
Marley 181 BL	6/23/2015	26.53	1.8	1.24
Marley 181-2 BL	6/23/2015	15.69	1.68	1.34
Marley 181-3 BL	6/23/2015	16.24	1.61	1.43
Marley 181 1YR	6/23/2015	31.88	1.97	1.15
Marley 178 BL	6/23/2015	25.31	1.72	1.08
Marley 178 1YR	6/23/2015	25.8	1.78	1.98
Marley 157-1 BL	6/23/2015	21.59	1.72	1.62
Marley 157-2 BL	6/23/2015	19.25	1.67	0.35
Marley 157 1YR	6/23/2015	18.75	1.71	2.35
Marley 186 BL	6/23/2015	16.81	1.56	0.81
Marley 186 1YR	6/23/2015	9.06	1.47	2.12
Marley 176-1 BL	6/23/2015	17.58	1.72	2.39
Marley 176-2 BL	6/23/2015	22.37	1.87	1.12
Marley 176-3 BL	6/23/2015	21.05	1.71	1.81
Marley 176 1YR	6/23/2015	24.46	1.85	2.31
Marley 211 BL	6/23/2015	25.77	1.9	0.56
Marley 35 BL	6/23/2015	15.92	1.49	1.63
Marley 35 1YR	6/23/2015	12.58	1.8	4.11

Marley 49 BL	6/23/2015	17.87	1.98	2.89
Marley 165 1YR	6/23/2015	29.07	1.99	2.15
Marley 193 BL	6/29/2015	27.82	1.85	0.91
Marley 193 1YR	6/29/2015	23.61	1.74	1.02
Marley 163 BL 1	6/29/2015	21.83	2.03	0.82
Marley 163 BL 2	6/29/2015	28.38	1.84	0.38
Marley 163 1YR	6/29/2015	13.85	2.06	1.09
Marley 162 BL	6/29/2015	16.3	1.94	0.92
Marley 162 1YR	6/29/2015	17.24	2	1.21
Marley 179 BL	6/29/2015	24.5	2.14	1.07
Marley 179 1YR	6/29/2015	29.63	2.08	1.17

Table 7. Concentrations of the 24 RNA libraries normalized to 4 nM for pooling and RNA sequencing.

Sample ID	ng/µl	Vol for 50 ng	Vol H2O	Index	Ave Size (bp)	Qubit Conc (ng/µl)	Conc (nM)	Vol HT1 for 4 nM (2 µl input)
186 BL	16.81	2.974	7.026	**2**	304	52	259.17	**127.58**
186 1YR	9.06	5.519	4.481	**7**	303	45.1	225.52	**110.76**
428 BL	75.61	1.322	18.678	**19**	308	41.3	203.17	**99.58**
428 1YR	76.54	1.306	18.694	**5**	304	34	169.46	**82.74**
163 BL	21.83	2.29	7.71	**6**	304	43	214.31	**105.16**
163 1YR	13.85	3.61	6.39	**15**	305	47.4	235.47	**115.74**
350 BL	85.14	1.174	18.826	**2**	305	50	248.39	**122.2**
350 1YR	78.85	1.268	18.732	**7**	296	52	266.18	**131.1**
176-3 BL	21.05	2.375	7.625	**19**	325	50	233.1	**114.56**
176 1YR	24.46	2.044	7.956	**5**	317	41.8	199.79	**97.9**
432 BL	65.23	1.534	18.466	**6**	309	39.7	194.67	**95.34**
432 1YR	41.88	1.194	8.806	**15**	318	38.7	184.39	**90.2**
179 BL	24.5	2.041	7.959	**2**	320	51	241.48	**118.74**
179 1YR	29.63	1.687	8.313	**7**	310	43.6	213.1	**104.56**
884 BL	18.13	2.758	7.242	**19**	307	40.7	200.87	**98.44**
884 1YR	9.22	5.423	4.577	**5**	304	49.9	248.7	**122.36**
181-3 BL	16.24	3.079	6.921	**6**	324	45.9	214.65	**105.32**
181 1YR	31.88	1.568	8.432	**15**	322	49.8	234.33	**115.16**
434 BL	32.61	1.533	8.467	**2**	310	40.5	197.95	**96.98**
434 1YR	94.2	1.062	18.938	**7**	305	33.6	166.92	**81.46**
193 BL	27.82	1.797	8.203	**19**	305	43.3	215.1	**105.56**
193 1YR	23.61	2.118	7.882	**5**	311	45.2	220.21	**108.1**
287 BL	60.53	1.652	18.348	**6**	312	34.8	169	**82.5**
287 1YR	79.37	1.26	18.74	**15**	325	56	261.07	**128.54**

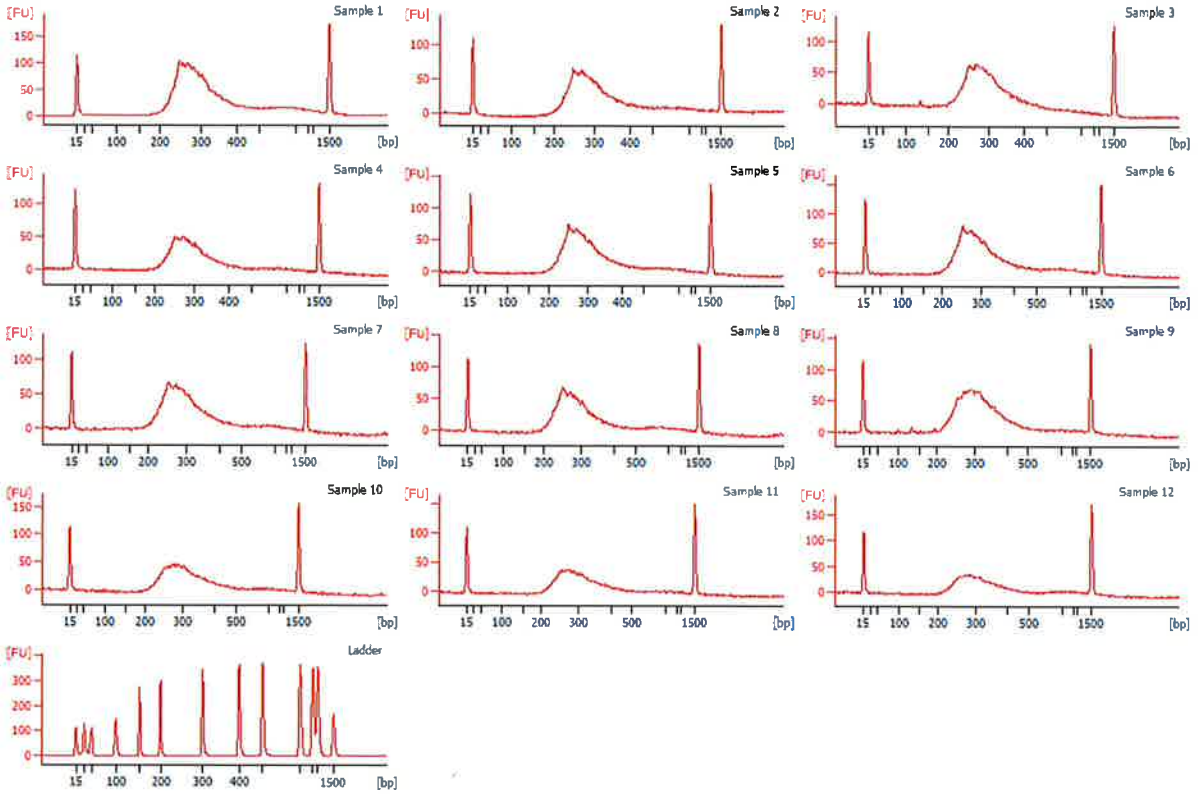

Figure 8. Bioanalyzer trace showing fragment size distribution for 12 RNA libraries used for RNA sequencing.

The Clustered flow cell was then run on the HiSeq 2000 using the paired end 200 cycle kit. Each lane of the 8 lane flow cell produced on average 237 million reads and the average % of the reads greater than Q30 for each lane was 86% (Figure 9).

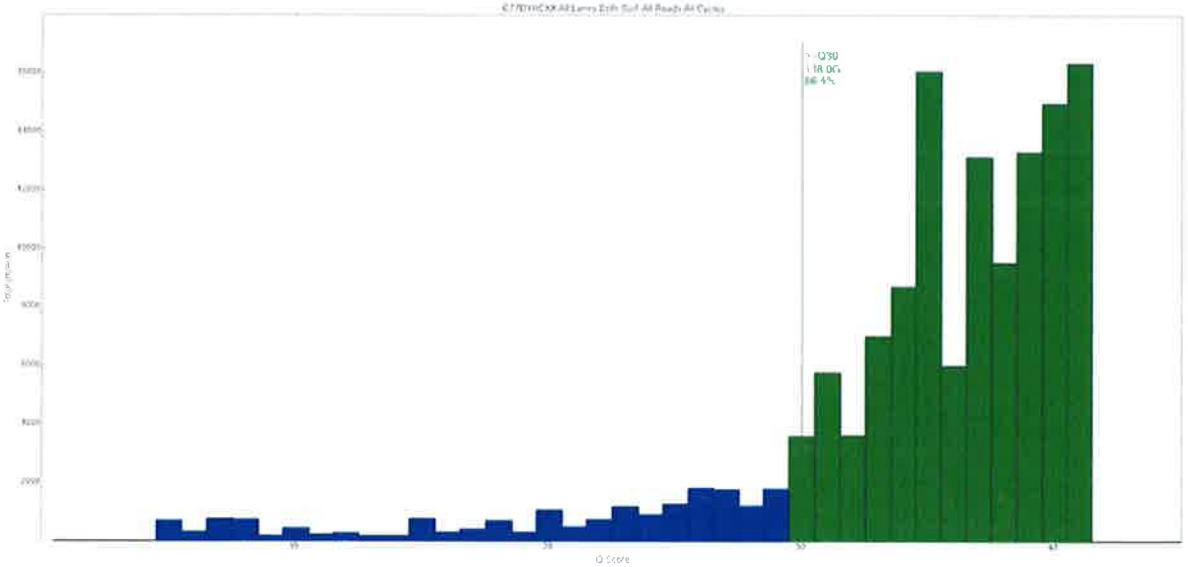

Figure 9. Bar graph showing the percentage of reads with a quality score of ≥Q30.

Task #3: Use whole transcriptome analysis in the CRC to examine expression of previously identified genes

During the year, no additional patient blood samples were collected. Total RNA was isolated from 257 peripheral blood samples from 106 patients (Table 8). RNA concentrations were 86.1±42.4 ng/μl (range 23.9-326.3 ng/μl), OD260/280 ratios were 2.16±0.09 (range 1.88-2.45), and RIN numbers were 8.37±0.40 (range 6.80-9.40). Eighty-four RNA samples were run on Affymetrix gene expression arrays with call rates of 59.76±1.28% (range 56.40-62.40%). Eighty-one baseline, 23 control waiting period complete, 62 intervention complete, and 91 six months after intervention time points were processed during this time period.

Table 8. Concentrations, purity measures, and call rates on gene expression arrays for RNA isolated from whole blood from CRC lifestyle participants.

Sample	Time Point	Concentration (ng/μl)	OD260/280	RIN	Call Rate (%)
C005	T2C	40.62	2.23	8.1	
C005	T3	77.70	2.18	8.3	
C005	T1	62.60	2.00	8.2	
C006	T1	75.91	2.18	8.2	
C006	T3	47.10	2.25	8.7	
C006	T2C	94.30	2.05	8.6	
C009	T3	68.01	2.12	8.5	
C009	T2C	119.29	2.23	8.7	
C015	T3	77.79	2.15	9.2	61.89
C015	T1	77.59	2.08	8.8	62.31
C015	T2V	66.33	2.15	9.0	60.67
C017	T3	52.47	2.27	8.4	
C017	T1	53.40	2.14	8.5	
C025	T3	51.59	2.26	8.8	61.03
C025	T2V	90.55	2.15	8.3	60.36
C025	T1	32.90	2.32	8.4	59.59
C026	T1	51.24	2.22	8.7	
C026	T2C	52.40	2.12	8.8	
C026	T3	95.60	1.88	8.4	
C036	T3	45.10	2.16	8.6	
C036	T1	112.40	2.07	8.4	
C036	T2V	75.20	2.09	8.6	
C039	T2V	57.70	2.13	8.7	
C039	T1	58.07	2.19	8.3	59.24

C039	T3	90.07	2.13	8.1	
C046	T1	126.41	2.10	8.2	59.25
C046	T3	114.30	2.10	8.0	
C050	T3	95.11	2.18	8.7	58.13
C050	T1	84.54	2.12	8.1	59.75
C050	T2V	67.95	2.08	8.1	59.48
C073	T1	26.04	2.00	8.8	60.15
C076	T2V	95.68	2.14	8.9	
C076	T3	72.35	2.13	9.1	
C076	T1	43.92	2.30	9.0	
C089	T2V	51.11	2.17	8.4	
C089	T1	103.90	2.06	8.4	
C089	T3	124.70	2.14	8.5	
C119	T2C	91.55	2.10	8.0	
C135	T2V	60.58	2.24	8.4	59.68
C135	T3	81.28	2.19	9.0	
C135	T1	110.00	2.16	8.7	
C163	T1	84.91	2.21	8.0	
C163	T2C	109.40	2.05	8.5	
C163	T3	52.40	2.27	8.1	
C165	T3	55.20	2.34	8.1	
C173	T1	43.69	2.11	8.7	
C173	T3	39.50	2.25	8.6	
C173	T2C	134.50	2.17	8.4	
C176	T2C	87.96	2.20	8.3	
C176	T1	129.50	2.17	8.4	
C182	T1	101.23	2.20	8.7	60.11
C182	T3	52.41	2.13	8.2	
C194	T3	57.70	2.26	8.5	60.16
C194	T2V	54.90	2.12	8.1	61.13
C194 (C477 T3)	T1	92.10	2.13	8.7	60.08
C195	T3	45.81	2.33	8.6	58.80
C195	T1	107.80	2.14	7.6	
C195	T2V	81.00	2.21	7.9	
C195	T1	107.80	2.14	7.6	
C200	T3	65.09	2.15	8.7	62.24
C200	T1	38.68	2.24	8.2	60.33
C200	T2V	73.71	2.38	9.2	61.56
C204	T1	123.00	2.16	8.4	59.26
C204	T2V	110.88	2.21	8.8	
C204	T3	50.36	2.11	8.7	
C206	T3	115.00	2.17	8.5	
C212	T2V	45.36	2.39	9.0	

C212	T3	56.50	2.05	8.8	
C215	T3	49.50	2.20	7.0	
C219	T1	67.63	2.12	9.2	60.93
C219	T3	100.04	2.08	8.8	58.90
C219	T2V	112.47	2.10	8.0	59.66
C241 (C457 T1)	T3	56.10	2.01	8.4	
C250	T1	73.43	2.16	8.9	60.38
C250	T2V	48.08	2.30	8.2	61.57
C250	T3	68.23	2.30	8.5	59.46
C251	T1	52.89	2.22	8.1	
C251	T2C	73.60	2.20	8.3	
C251	T3	60.20	2.15	8.1	
C262	T3	143.51	2.19	8.1	57.40
C262	T1	77.55	2.12	8.4	58.74
C262	T2V	72.89	2.19	8.6	
C302	T1	185.07	2.10	8.0	57.71
C303	T2V	89.79	2.06	8.5	
C303	T3	63.10	2.20	---	
C323	T1	64.58	2.36	8.2	59.88
C323	T2V	133.58	2.29	8.4	59.28
C323	T3	60.00	2.26	8.8	60.43
C326	T1	98.58	2.25	8.7	56.40
C326	T3	194.69	2.13	8.5	
C326	T2V	156.69	2.17	8.9	
C333	T2V	89.15	2.22	8.1	
C333	T1	128.50	2.16	8.1	
C333	T3	78.80	2.15	8.1	
C360	T3	78.16	2.17	7.4	
C360	T1	58.10	2.23	8.1	
C360	T2V	56.00	2.14	8.2	
C366	T1	69.50	2.19	8.1	
C366	T2C	74.70	2.12	8.4	
C366	T3	66.60	2.10	8.0	
C375	T1	49.78	2.21	8.1	
C375	T3	101.90	2.15	8.0	
C375	T2C	44.10	2.02	8.0	
C387 (C926 T1)	T3	112.00	2.18	8.6	
C412	T1	119.01	2.21	8.3	59.68
C412	T2V	59.33	2.13	9.2	
C412	T3	147.74	2.15	8.5	
C414	T2C	62.58	2.07	8.8	
C414	T3	54.80	2.27	8.2	
C424	T2C	104.49	2.14	8.0	

C424	T2C	104.49	2.14	8.0	
C424	T1	84.00	2.17	8.1	
C424	T3	71.70	2.05	8.1	
C439	T1	48.38	2.17	8.4	
C439	T2V	60.30	2.13	8.6	
C439	T3	41.70	2.12	8.7	
C455	T1	124.06	2.11	8.3	
C455	T2C	117.19	2.10	8.3	
C455	T3	54.50	2.29	7.9	
C471	T2V	150.62	2.21	8.0	58.97
C471	T1	58.76	2.24	8.5	
C471	T3	56.99	2.12	8.2	
C477	T1	77.30	2.09	8.2	
C477	T2C	44.70	2.14	9.1	
C481	T3	35.24	2.29	8.6	
C481	T1	129.23	2.11	8.0	59.40
C481	T2V	129.23	2.11	8.0	
C496	T1	58.37	2.11	8.9	61.02
C496	T2C	135.80	2.08	8.2	
C496	T3	158.41	2.20	8.5	
C498 (C822 T1)	T3	56.00	2.24	8.0	
C500 (C725 T3)	T1	82.20	2.05	8.3	
C500	T2V	52.60	2.17	8.6	
C500	T3	104.00	2.15	8.3	
C521	T1	90.46	2.26	8.6	61.13
C521	T2V	172.36	2.16	6.8	61.19
C521	T3	64.83	2.15	8.8	
C525	T3	60.27	2.26	8.3	
C525	T2C	126.66	2.12	7.9	
C530	T1	51.88	2.22	8.6	60.72
C530	T2V	54.16	2.44	8.6	60.65
C530	T3	29.20	2.16	8.9	59.34
C532	T2V	58.50	2.24	7.9	
C533	T3	115.75	2.08	8.2	60.95
C534	T2V	98.87	2.23	8.6	60.50
C534	T1	163.14	2.19	7.9	58.14
C534	T3	64.05	2.22	8.8	
C550	T1	52.30	2.17	8.3	
C550	T2V	112.70	2.10	7.4	
C550	T3	60.80	2.16	8.5	
C578	T1	65.40	2.13	8.6	
C595	T1	133.00	2.17	8.4	
C595	T2V	40.30	2.22	8.7	

C595	T3	55.70	2.09	8.7	
C604	T1	88.39	2.13	8.3	
C604	T2V	36.38	2.19	8.2	
C604	T3	72.63	2.25	8.8	
C608	T1	96.13	2.29	8.4	59.42
C608	T2V	174.90	2.21	8.4	59.04
C608	T3	69.20	2.07	7.8	60.10
C614	T2C	123.23	2.13	7.9	
C614	T3	48.40	2.27	7.5	
C624	T1	35.24	2.29	8.6	62.40
C639	T1	114.74	2.09	8.6	
C639	T2C	51.80	2.27	7.6	
C639	T3	57.50	2.15	8.3	
C641	T2V	92.46	2.24	8.8	57.01
C641	T3	181.54	2.22	7.3	56.64
C641	T1	94.69	2.27	8.6	61.46
C645	T3	308.49	2.09	8.3	
C645	T2V	103.21	2.19	8.6	
C645	T1	209.90	2.17	8.4	
C649	T2V	49.46	2.30	8.4	59.78
C649	T1	88.97	2.13	8.2	60.62
C649	T3	29.73	2.03	9.4	
C669	T2V	91.57	2.44	8.1	60.71
C669	T3	64.50	2.12	8.0	58.68
C672	T2V	108.90	2.17	8.4	
C672	T3	89.40	2.07	8.6	
C676	T1	133.78	2.11	7.8	59.69
C676	T2V	85.82	2.12	7.9	60.08
C676	T3	127.03	2.12	7.0	
C683	T1	198.51	2.15	8.5	58.79
C683	T3	90.48	2.13	8.7	
C683	T2V	53.09	2.16	8.3	
C687	T2V	124.80	2.21	8.7	57.38
C687	T3	42.91	2.16	8.7	
C687	T1	89.94	2.08	8.0	
C698	T2V	71.70	2.27	8.8	60.85
C698	T1	126.59	2.14	8.7	
C698	T3	57.51	2.04	9.0	
C705	T2V	119.50	2.13	8.6	58.25
C705	T1	95.73	2.07	8.4	60.29
C705	T3	326.25	2.16	---	
C718	T1	63.60	2.09	8.6	
C735	T3	47.30	2.10	8.1	

C735	T2V	50.20	2.29	7.8	
C737	T2C	49.20	2.45	7.5	
C751	T3	141.62	2.17	8.1	58.92
C751	T2V	78.84	2.20	8.7	
C751	T1	171.94	2.17	8.4	58.92
C754	T3	157.32	2.16	7.4	
C754	T1	58.90	2.15	8.2	
C754	T2C	76.30	2.11	8.6	
C755	T1	58.06	2.09	8.7	57.76
C755	T2V	48.65	2.04	9.0	
C755	T3	23.85	2.17	8.5	
C767	T1	114.22	2.12	7.9	60.21
C772 (C706 T1)	T3	30.40	2.16	8.6	
C777	T1	53.50	1.91	8.0	
C777	T2C	82.10	2.06	8.6	
C777	T3	45.80	2.18	8.3	
C778	T2V	106.87	2.21	8.4	58.71
C778	T1	91.89	2.11	8.5	
C778	T3	74.37	2.14	8.0	
C792	T2V	72.20	2.13	8.6	
C792	T3	39.70	1.99	8.1	
C811	T1	114.40	2.14	7.6	
C811	T2V	152.30	2.14	7.9	
C811	T3	81.50	2.19	7.9	
C814	T2V	114.07	2.11	8.8	59.23
C814	T1	107.87	2.07	8.4	
C814	T3	88.60	1.92	8.2	60.08
C815	T3	74.50	2.13	8.4	
C822	T2V	46.70	2.14	8.1	
C850	T3	175.47	2.18	7.8	
C850	T1	121.41	2.14	7.7	57.99
C862	T1	134.57	2.11	8.2	
C862	T2V	121.63	2.10	8.4	
C862	T3	55.70	2.17	8.0	
C870	T1	51.71	1.94	8.6	
C870	T2V	51.32	2.24	8.2	
C870	T3	68.63	2.17	8.6	
C870	T2V	51.32	2.02	8.2	
C870	T3	68.63	2.17	8.6	
C870	T1	51.71	2.04	8.5	
C882	T3	49.79	2.17	8.7	62.09
C882	T2V	47.89	2.35	8.5	61.47
C882	T1	37.56	2.10	8.5	

C887	T1	38.37	2.15	8.5	
C887	T2V	53.50	2.04	8.5	59.35
C887	T3	53.50	2.08	8.7	
C898	T1	149.73	2.17	8.5	
C898	T3	93.83	2.10	8.5	
C898	T2V	113.73	2.16	8.3	
C919	T3	122.09	2.16	8.3	59.81
C919	T1	86.76	2.21	8.5	59.68
C919	T2V	60.78	2.15	8.8	
C932	T1	111.89	2.27	8.5	59.97
C932	T3	105.10	2.15	8.3	58.32
C932	T2V	141.50	2.17	8.6	60.31
C940 (C881 T1)	T3	65.10	2.18	8.5	
C949	T2V	190.11	2.11	8.8	
C962	T1	52.01	2.00	8.0	60.17
C989	T2V	115.62	2.27	8.7	58.29
C989	T1	78.41	2.25	8.8	59.49
C989	T3	101.35	2.23	8.6	60.03

Task #4: Investigate gender and patient subgroup differences in molecular response

Nothing to report.

Task #5: Discover new genetic influences on heart disease by profiling micro-RNAs and rare RNA transcripts

Tasks #3, #4, and #5 are using current patients in the Cardiovascular Risk Clinic. No new participants entered the CRC program and no new blood samples were collected during the year.

Task #6: Develop systems biology approach to integrate various types of risk factor data

Task #6 will utilize the large-scale DNA and RNA sequence data generated in Tasks #1 – #5, along with other CVD risk factor data collected in the Integrative Cardiac Health Program. To derive maximum information from the data, we are collaborating with scientists who have expertise in systems biology to integrate all of the different types of data. This approach will allow us to uncover inter-relationships and patterns within the data that may not be apparent when each modality is analyzed independently. Progress on this task will be made when we have sufficient DNA or RNA sequence data for analysis.

Discussion

During the year, we made progress in examining patterns of DNA methylation across the genome in circulating leukocytes in response to cardiovascular risk reduction

(lifestyle and surgically-assisted). DNA was isolated from 102 whole blood samples in the following groups: intensive lifestyle baseline (n=23) and one year (n=23), laparoscopically placed adjustable gastric banding (LAGB) baseline (n=23) and one year (n=23), and control baseline (n=5) and one year (n=5). Methyl-Mini Sequencing has now been completed on a total of 60 samples, 30 samples are currently undergoing sequencing, and 30 additional samples will be sequenced as soon as possible.

Problems encountered during the year and plans to resolve them include:

The HiSeq, MiSeq, and cBot machines had service contract coverage throughout the year and all are functioning normally.

The logistical issue of getting supplies continued throughout much of the year. We hope this issue can be resolved so that supplies will be ordered and delivered in a timely manner.

4. KEY RESEARCH ACCOMPLISHMENTS:

1. Completed 60 samples for genome-wide methylation analysis and began interpretation of results
2. Completed 72 samples for large-scale RNA sequencing

5. CONCLUSION: During the next year, our main focus will be on keeping all machines running and generating as much quality RNA sequencing and genome-wide DNA methylation data as possible. Once we have sufficient preliminary data, we will begin integrating the various types of data and begin preparing abstracts for presentation and publication.

6. PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS:

Publications

1. Blackburn HL, McErlean S, Jellema GL, van Laar R, Vernalis MN, Ellsworth DL. Gene expression profiling during intensive cardiovascular lifestyle modification: Relationships with vascular function and weight loss. Genomics Data 2015;4:50-53.
2. Ellsworth DL, Mamula KA, Blackburn HL, McDyer FA, Jellema GL, van Laar R, Costantino NS, Engler RJ, Vernalis MN. Importance of substantial weight loss for altering gene expression during cardiovascular lifestyle modification. Obesity 2015;23:1312-1319.

7. INVENTIONS, PATENTS AND LICENSES: Nothing to report.

8. REPORTABLE OUTCOMES: Nothing to report.

9. OTHER ACHIEVEMENTS: Nothing to report.

10. REFERENCES: Nothing to report.

11. APPENDICES: Nothing to report.

The report does make it clear the research team has done an extensive amount of work. However it is not always clear where exactly the project is in terms of its goals as stated in the SOW. Therefore, please resubmit the report with the following additions:

1. For each figure or table please provide a title and a legend.
Table titles and Figure legends have been added to the report.

2. Where there are anomalous data, e.g. some negative values for concentration and OD in the table starting on page 19, please include an explanation for how the data/sample will be handled.
Samples with very low concentrations or those with nonsense (negative) concentration values were treated as 0 and were not selected for RNA sequencing. These samples were used in their entirety for previous research. Any remaining RNA was below the level of detection.

3. The report provides information on the number of patients recruited and samples acquired, but not always on the number of patients in each cohort. In Task 1 this information is provided and is very helpful (e.g. 23 LAGB baseline). Please provide this information for all the tasks.

a. Task 2, how many adipose tissue samples and how many peripheral blood samples and at what time points (e.g. at time of surgery (baseline), 1yr after etc.) were acquired.
Task #2: During the year, no new patients undergoing laparoscopically placed adjustable gastric banding (LAGB) were enrolled in the study. No additional follow-up blood samples or adipose tissue samples were collected. A summary of time points for which RNA was isolated from peripheral blood during the year is as follows: baseline pre-surgery (n=14), five to seven months post-surgery (n=24), one year post-surgery (n=10), one year one month to one year eleven months (n=13), two years to two years eleven months (n=2), three years six months (n=1), four years six months (n=1), five years (n=3), five years one month to five years eleven months (n=7), and six years post-surgery (n=4).

b. Task 3, please provide a breakdown of samples acquired by cohort, e.g. x samples from y patients at initiation of CRC lifestyle program; w samples from z patients at 1 year etc.
During the year, no additional patient blood samples were collected. Total RNA was isolated from 257 peripheral blood samples from 106 patients. Eighty-one baseline, 23 control waiting period complete, 62 intervention complete, and 91 six months after intervention time points were processed during the year.

Note that the samples collected and analyzed for Tasks #2 and #3 are also used for Tasks #4 and #5.

4. For Task 6, please include an estimation of how much DNA or RNA sequence data you anticipate needing before beginning this task and when you anticipate this milestone will be achieved.
In the time since this annual report was submitted (July 2015), we have made substantial progress in collecting and processing DNA and RNA data for many of the tasks. For Task #1, Reduced Repression Bisulfite Sequencing for detecting genome-wide patterns of DNA methylation has been completed for all samples. Comparative analyses are in progress. The RNA sequencing aspect of Task #2 reached a bench mark for data collection in December of 2015. All LAGB patients that have a blood sample from baseline and the one-year time point and an age- and gender-matched patient in the healthy style intervention program have complete RNA sequence data. Sufficient RNA sequence data has been collected to begin processing and analysis. These patients were prioritized for sequencing because it will allow us to execute several comparative studies for Tasks #3-5 without having to complete all of the data collection.

Tasks #3-5 are still ongoing, but sufficient data has been collected to perform some analyses. RNA sequencing on all samples is expected to be completed by May 31, 2016. The DNA methylation and RNA sequencing data require unique skills and methods for processing, analysis, and interpretation. One limitation for processing and analyzing RNA sequence data is the very large amounts of data generated, as well as the sizes of the sequence files. These analyses require substantial computing power and processing and analysis takes several weeks or more of computing time to analyze one sequence run. Total completion of Task 6 is anticipated by the completion of this award.